

UNIVERSIDAD  
NACIONAL  
DE TUCUMÁN

2-Hidroinformática y tecnologías para la gestión del agua.

## Hacia un Sistema de Pronóstico en Tiempo Real de Niveles de Marea en el Río de la Plata Enriquecido con Herramientas de Aprendizaje Automático

Diego Silva Piedra <sup>a</sup>, Ianla Rivera <sup>b</sup>, Alejo Pilosio <sup>a</sup>, Pablo Ezzati <sup>a</sup>, Mónica Fossati <sup>b</sup>

<sup>a</sup> Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay, [diegos@fing.edu.uy](mailto:diegos@fing.edu.uy), [apilosio@fing.edu.uy](mailto:apilosio@fing.edu.uy), [pezzatti@fing.edu.uy](mailto:pezzatti@fing.edu.uy).

<sup>b</sup> Instituto de Mecánica de los Fluidos e Ingeniería Ambiental, Facultad de Ingeniería, Universidad de la República, Uruguay, [irivera@fing.edu.uy](mailto:irivera@fing.edu.uy), [mfossati@fing.edu.uy](mailto:mfossati@fing.edu.uy).

### RESUMEN:

PronUy\_RPFM es una herramienta diseñada para predecir el nivel de la marea en el Río de la Plata y su Frente Marítimo de manera operativa. Fue creado por investigadores de los institutos de Mecánica de los Fluidos e Ingeniería Ambiental (IMFIA) y de Computación (INCO), de la Facultad de Ingeniería (FIng) de la Universidad de la República (UdelaR) Uruguay. La herramienta ofrece el pronóstico con datos horarios del nivel de mar para los próximos tres días de manera pública en la web [www.fing.edu.uy/imfia/pronostico-marea](http://www.fing.edu.uy/imfia/pronostico-marea). PronUy\_RPFM se basa en el modelo numérico en elementos finitos TELEM2D. En este trabajo se presentan los avances en la aplicación de métodos de aprendizaje automático (AA) en la herramienta PronUy\_RPFM en la etapa de pronóstico, revisando su efectividad según la comparación con datos medidos. Concretamente, se presenta la implementación y resultados del marco LightGBM para pronóstico de niveles de marea. Los modelos fueron entrenados con datos de nivel medidos por mareógrafos a lo largo de la costa uruguaya y argentina.

PALABRAS CLAVE: pronóstico, niveles de marea, aprendizaje automático, Río de la Plata.

### 1 INTRODUCCIÓN

El Río de la Plata es uno de los estuarios más extensos del mundo, ubicado entre Argentina y Uruguay, con un caudal medio de 26.000 m<sup>3</sup>/s, principalmente de los ríos Paraná y Uruguay. Su compleja dinámica de mareas meteorológicas requiere herramientas de pronóstico para optimizar la navegación, gestión portuaria, emergencias y operaciones costeras.

En el IMFIA en conjunto con el INCO de la FIng de la UdelaR de Uruguay se ha mantenido una línea de investigación continua centrada en la dinámica hidrodinámica y las propiedades del agua del Río de la Plata<sup>[1],[2]</sup>. Particularmente, se ha implementado el sistema PronUy\_RPFM, un desarrollo automatizado para el pronóstico operativo de condiciones meteo-oceánicas en el Río de la Plata y su Frente Marítimo<sup>[3]</sup>.



Eventos extremos climáticos  
y sociales en el espacio AUGM.  
Lo urgente y lo importante en el  
nexo agua-energía-ambiente



UNIVERSIDAD  
NACIONAL  
DE TUCUMÁN

Aunque los modelos hidrodinámicos han sido tradicionales para pronósticos de marea, los modelos de AA han ganado relevancia por su desempeño en series temporales. LightGBM ofrece perspectivas innovadoras como herramienta complementaria, con aplicaciones exitosas en diversas áreas y antecedentes prometedores en predicción de niveles de agua<sup>[4]</sup>.

El objetivo de este trabajo es evaluar la implementación del algoritmo de AA LightGBM como herramienta complementaria para el pronóstico de niveles de marea en el Río de la Plata, analizando su desempeño en diferentes estaciones (Isla de Flores -IdF-, Buenos Aires -BsAs- y Montevideo -MVD-) y horizontes de predicción (6, 12, 24, 48 y 72 horas), con vistas a su futura integración en el sistema operativo PronUy\_RPFM.

## 2 DESARROLLO DEL TRABAJO

### 2.1 Datos medidos

El conjunto de datos empleado en este estudio proviene de los registros de nivel de marea obtenidos mediante la red de mareógrafos instalados a lo largo de la costa uruguaya por la Administración Nacional de Puertos (ANP) y a lo largo de la costa argentina por el Servicio de Hidrografía Naval (SHN), que reportan valores con una frecuencia de cinco minutos. La serie temporal cubre el período 2017–2025, con diferentes longitudes y fechas de inicio según la estación. Se consideraron en este trabajo las mediciones de MVD, BsAs e IdF, preprocesadas para derivar promedios horarios, remover valores atípicos e interpolar linealmente pequeños grupos de valores faltantes. Como se observa en la Figura 1, las estaciones presentan diferencias tanto en la extensión temporal como en la continuidad de las series, con intervalos de registros incompletos (para facilitar la comparación visual, los niveles se representan con un desplazamiento vertical arbitrario).

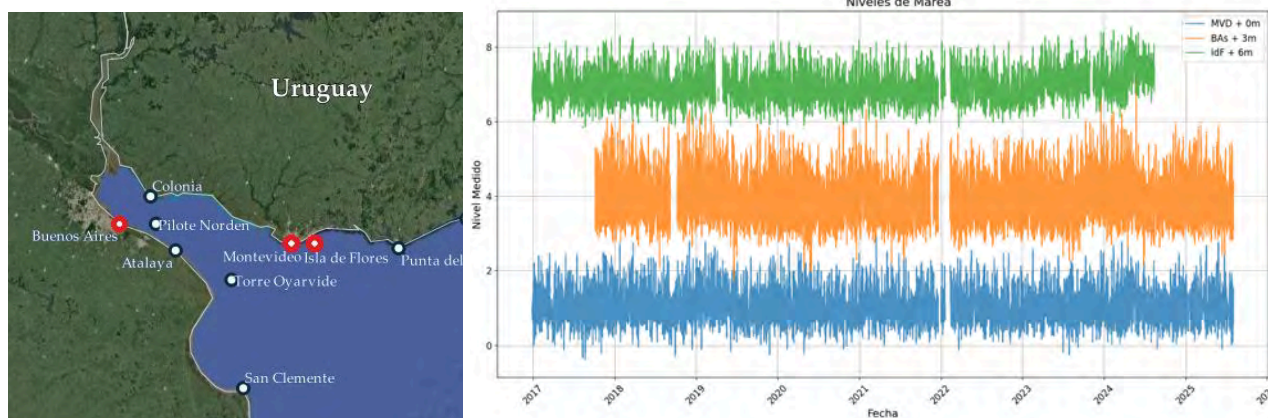
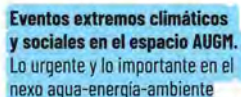


Figura 1. Ubicación de mareógrafos (izq.), y registros horarios de nivel de marea registrados en MVD (azul), BsAs (anaranjado) e IdF (verde).

#### 2.1.1 El Río de la Plata

La dinámica del Río de la Plata es compleja, influenciada por su geomorfología y forzantes como los caudales fluviales, los vientos fuertes y presiones atmosféricas, y las mareas astronómicas y meteorológicas provenientes del océano. Es un estuario mayormente somero, con profundidades que no superan los 10m hasta la sección MVD-Punta Piedras (Figura 1), y superando los 20m de profundidad en la zona exterior. La marea meteorológica es particularmente relevante, con su onda



ingresando por la plataforma argentina y sufriendo transformaciones por los vientos, la presión, y la fricción de fondo. Los niveles máximos se asocian a vientos del sudeste, y los mínimos a vientos del oeste-noroeste<sup>[3][5]</sup>.

## 2.2 Descripción del Modelo

### 2.2.1 Light Gradient Boosting Machine

LightGBM (Light Gradient Boosting Machine) es un algoritmo de AA de Microsoft que construye iterativamente un ensamble de árboles de decisión mediante gradient boosting<sup>[6]</sup>. Su fortaleza principal es capturar automáticamente relaciones no lineales complejas sin especificaciones previas sobre la forma funcional. Su alta eficiencia computacional y capacidad para grandes volúmenes de datos lo hacen robusto para modelado predictivo de series temporales ambientales en sistemas complejos como estuarios.

Los árboles de decisión incluyen un nodo raíz (entrada de datos), nodos internos (reglas de división) y nodos hoja (predicciones finales). Utilizando niveles de marea históricos como entrada, el modelo puede generar predicciones mediante subdivisiones basadas en valores de corte aprendidos<sup>[7]</sup>. LightGBM implementa gradient boosting, creando un modelo ensamblado que combina predicciones individuales para lograr mayor precisión y robustez predictiva. Asumiendo que  $X$  es la entrada, la salida final  $Y_I$  del modelo se obtiene como:

$$Y_I = \sum_{i=1}^I T_i(X, \theta_i) \quad (1)$$

donde  $i$  es el índice del árbol,  $I$  es el número total de árboles,  $Y_I$  es la salida resultante de la suma de los  $I$  árboles,  $\theta_i$  representa la estructura topológica, y  $T_i$  es la salida del  $i$ -ésimo árbol. En el proceso de aprendizaje, se debe especificar una función objetivo para cada árbol:

$$Obj_i = L(y - y_i) + \Omega_i \quad (2)$$

$Obj_i$  es la función objetivo del  $i$ -ésimo árbol, y  $L$  es la función de pérdida que se utiliza para evaluar los errores de los resultados del modelo en relación con las salidas objetivo.  $\Omega_i$  es una función de regularización para mejorar la robustez del modelo y prevenir el sobreajuste.

### 2.2.2 PronUy\_LightGBM para pronóstico de niveles en el Río de la Plata

Este modelo utiliza inputs creados a partir de los niveles horarios de agua históricos para predecir los niveles de agua subsiguientes. Considerando que LightGBM tiene como salida un único valor, se desarrolló *PronUy\_LightGBM*, que construye un único modelo de regresión y emplea un enfoque recursivo para generar salidas multidimensionales. En este proceso recursivo, cada salida obtenida (predicción del nivel de la hora siguiente) por el modelo se utiliza como una de las entradas para el siguiente paso de predicción, permitiendo derivar salidas L-dimensionales mediante una recursión continua. En la Figura 2 se representa este proceso.

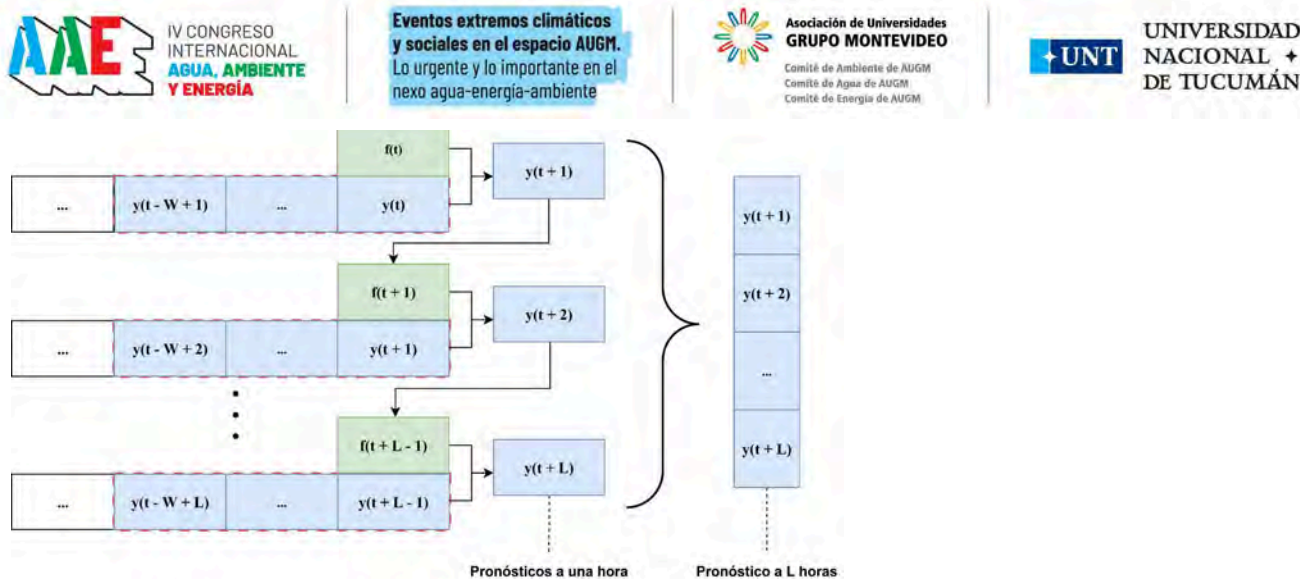


Figura 2. Diagrama de entradas y salidas de Pronuy\_LightGBM.

Donde  $y(t)$  y  $f(t)$  representan el nivel de marea en el tiempo  $t$  y el resto de las características de entrada para predecir el nivel en  $t + 1$ , respectivamente;  $W$  es el tamaño de la ventana temporal de niveles anteriores observados; y  $L$  es el tamaño del pronóstico deseado.  $f(t)$  incluye características básicas, como hora, minuto y día; y características calculadas a partir de datos de nivel, como media móvil y desviación estándar móvil.

Tanto  $W$  como los hiperparámetros del modelo fueron optimizados mediante validación. Los datos se dividieron en conjuntos de datos de entrenamiento y prueba. Durante el proceso de entrenamiento, se utilizó 20% de los datos de entrenamiento para comparar los modelos entrenados utilizando diferentes valores de parámetros (validación). Se utilizaron los datos de prueba para evaluar la precisión del modelo. Según disponibilidad de datos, en MVD y BsAs se utilizó el año 2025 como conjunto de prueba, y en IdF, el 2024. En cada caso se pronosticó cada día del conjunto de prueba para distintos largos de predicción  $L$  y se registró el RMSE.

### 2.3 Resultados

Para evaluar la precisión del modelo, se asignó el último año de datos como conjunto de prueba. Por lo tanto, en cada caso, se utilizó el primero de enero de 2025 o 2024 como índice de fraccionamiento entre el conjunto de entrenamiento y el de prueba. En la Figura 3 se ve la evolución del RMSE en función de  $L$  para las distintas estaciones. Para 6, 12, 24, 48 y 72 horas, los errores reportados promedio siguen una tendencia parecida, aumentando el error al incrementar el largo de predicción. Los errores mínimos se dan con la predicción a 6 horas en todas las estaciones. En términos absolutos, IdF registró los menores errores con valores entre 0.130 m (6h) y 0.168 m (72h), mientras que BsAs presentó los mayores errores absolutos con un rango de 0.263 m (6h) a 0.310 m (48h), y MVD mostró valores intermedios entre 0.170 m y 0.225 m. Al normalizar los errores por la variabilidad característica de cada estación para realizar una comparación justa entre ubicaciones con diferentes amplitudes de marea, IdF mantuvo consistentemente el mejor desempeño relativo; BsAs presentó los mayores errores relativos; y MVD ocupó una posición intermedia acercándose más al desempeño de BsAs, manteniendo lo observado en términos absolutos.

Se cuantificó la relevancia de las entradas en la capa inicial mediante el análisis de sus contribuciones. Este indicador se calcula como la proporción entre la frecuencia con la que *PronUy\_LightGBM* selecciona cada entrada para la división de nodos en los árboles y el número total de selecciones realizadas durante el proceso de modelado.



Eventos extremos climáticos y sociales en el espacio AUGM. Lo urgente y lo importante en el nexo agua-energía-ambiente



Comparación de RMSE Promedio: Original vs Normalizado

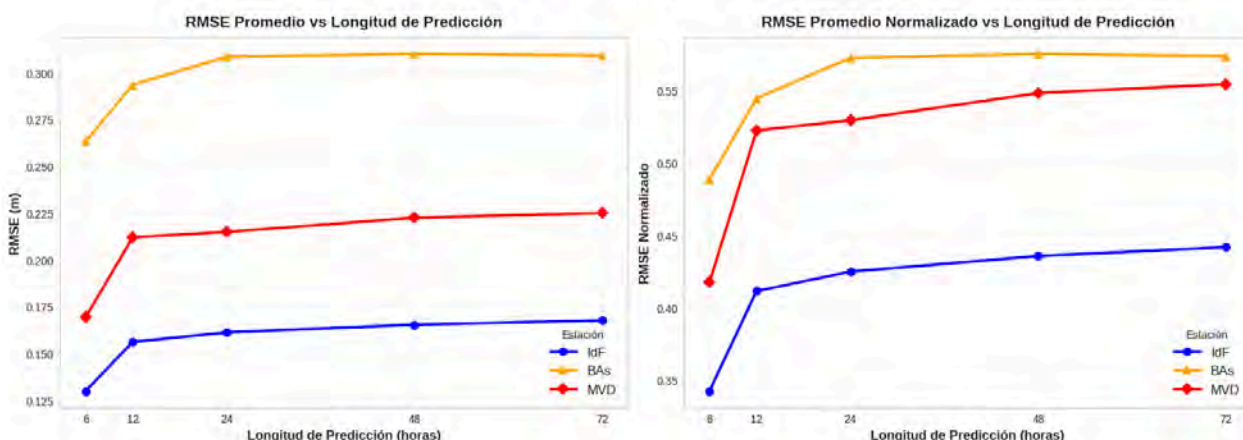


Figura 3. Variación en el RMSE promedio y contornos de percentil 1-99 en función de la longitud de predicción para las distintas estaciones.

Las 5 entradas más importantes de cada estación se muestran en la Figura 4. El nivel de agua en el paso temporal inmediatamente anterior ( $y_{-1}$ ,  $y(t)$ ) demostró ser consistentemente la variable de mayor influencia en las tres estaciones, siendo el factor más determinante para predecir el nivel en el paso siguiente.

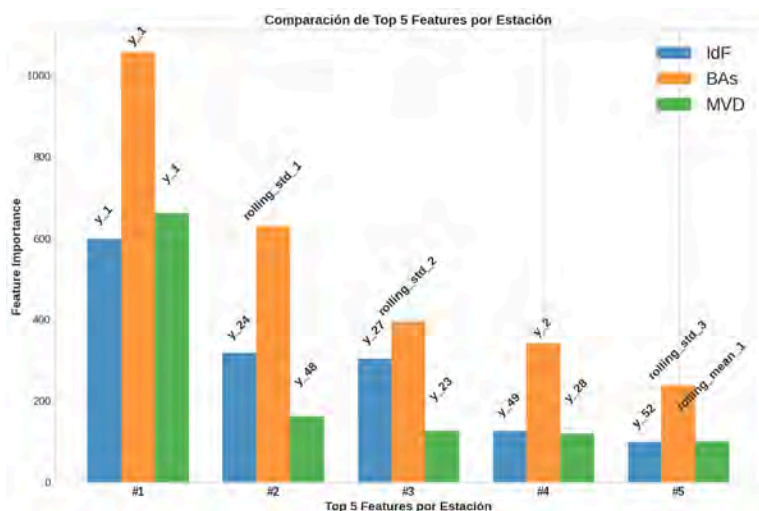


Figura 4. Características de entrada más incidentes para las distintas estaciones.

En IdF y MVD, los niveles registrados con 24, 27, y 48 horas de antelación ( $y_{24}$ ,  $y_{27}$ ,  $y_{28}$ ) también mostraron una contribución significativa. Sin embargo, BsAs exhibe un patrón diferente, donde las características de desviación estándar móvil ( $rolling\_std\_n$ ) ocupan posiciones prominentes en el top 5, sugiriendo que la variabilidad del nivel de agua es un factor crítico para la predicción en esta estación.

### 3 CONCLUSIONES

Se presentaron los primeros avances de integración de AA en PronUy\_RPFM para el pronóstico de niveles mediante *PronUy\_LightGBM*, que mostró un desempeño competitivo, con errores que



IV CONGRESO  
INTERNACIONAL  
AGUA, AMBIENTE  
Y ENERGÍA

Eventos extremos climáticos  
y sociales en el espacio AUGM.  
Lo urgente y lo importante en el  
nexo agua-energía-ambiente



Asociación de Universidades  
GRUPO MONTEVIDEO  
Comité de Ambiente de AUGM  
Comité de Agua de AUGM  
Comité de Energía de AUGM



UNIVERSIDAD  
NACIONAL  
DE TUCUMÁN

aumentan de forma consistente con el horizonte de predicción, pero manteniéndose en valores razonables para aplicaciones prácticas. La estación de IdF registró los mejores resultados con un RMSE mínimo de 0,130m a 6 horas, mientras que BsAs presentó los mayores errores promedio con un RMSE máximo de 0,310m a 48 horas, evidenciando que las características hidrodinámicas locales y calidad de datos pueden ser tan determinantes como su volumen. La normalización de errores confirmó que IdF presenta la mejor capacidad predictiva tanto en términos absolutos como relativos, independientemente de las diferencias en amplitud de marea entre estaciones.

El análisis de importancia de variables confirmó el carácter auto-regresivo de las series, siendo el nivel inmediatamente anterior el predictor más relevante en todas las estaciones. Sin embargo, se identificaron patrones diferenciados entre ubicaciones: en IdF y MVD la significativa contribución de niveles con retardos de 24 y 48 hs evidencia la fuerte influencia de las componentes de marea diurnas y semidiurnas<sup>[2]</sup>; BsAs presenta dominancia de características de variabilidad (*rolling\_std*) dominan entre las más importantes, sugiriendo que la irregularidad del nivel es crítica para su predicción, posiblemente debido a condiciones hidrodinámicas particulares. Esta variabilidad indica que el modelo se adapta a características específicas de cada ubicación.

Los resultados obtenidos constituyen una prueba de concepto alentadora para la incorporación de técnicas de inteligencia artificial al sistema PronUy\_RPFM, aportando una alternativa eficiente en términos computacionales. A futuro, se prevé explorar enfoques híbridos que combinen modelos hidrodinámicos y de AA, así como evaluar arquitecturas más avanzadas (e.g., redes neuronales recurrentes o modelos de atención) y explorar la inclusión de variables meteorológicas explícitas (viento, presión atmosférica) como entradas al modelo.

#### 4 AGRADECIMIENTOS

Este trabajo cuenta con el apoyo económico del proyecto ANII FMV “Sistema de pronóstico en tiempo real de niveles de marea en el Río de la Plata enriquecido con asimilación de datos”

#### 5 REFERENCIAS

- [1] Santoro, P., Fossati, M., & Piedra Cueva, I. (2013). Study of the meteorological tide in the Río de la Plata. *Continental Shelf Research*, 60, 51-63. 10.1016/j.csr.2013.04.018
- [2] Fossati M., Santoro P., Mosquera R., Martinez C., Ghiardo F., Ezzatti P., Pedocchi F., y Piedra-Cueva I. Dinámica de flujo, del campo salino y de los sedimentos finos en el Río de la Plata. *RIBAGUA - Revista Iberoamericana del Agua*, 1:48–63, 2014.
- [3] Fossati M., Balparda D., Sellanes L., Silva D., Jackson M., y Ezzatti P. Desarrollo del sistema de pronóstico del río de la plata y su frente marítimo: Pronuy\_rpfm. *Ribagua*, 9(2):25–40, 2022. doi:10.1080/23863781.2023.2210262.
- [4] Gan, M., Pan, H., Chen, Y., Pan, S., 2021a. Application of the variational mode decomposition (VMD) method to river tides. *Estuar., Coast. Shelf Sci.* 261 <https://doi.org/10.1016/j.ecss.2021.107570>
- [5] Sellanes, L., Fossati, M., Silva, D., & Ezzatti, P. (2023). Mejoras en la Implementación del Modelo Numérico base del Pronóstico de Niveles del Río de la Plata y Frente Marítimo: PRONUy\_RPFM. *Mecánica Computacional Vol XL* (págs. 1681-1690).
- [6] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., y Liu T.-Y. LightGBM: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30:3146–3154, 2017. doi:10.5555/3294996.3295074.
- [7] Gan, M. et al. (2024). An improved machine learning-based model to predict estuarine water levels. *Ocean Modelling*, 190, 102376. doi: 10.1016/j.ocemod.2024.102376.